Business Experimentation and Causal Methods

Prof. Fradkin

Topic: Statistics Refresher and Randomized Assignment of Treatment

What we know already

1. Example of a causal problem (in-person class vs zoom)
2. Potential outcomes
3. Average treatment effects
4. Selection bias

This Time

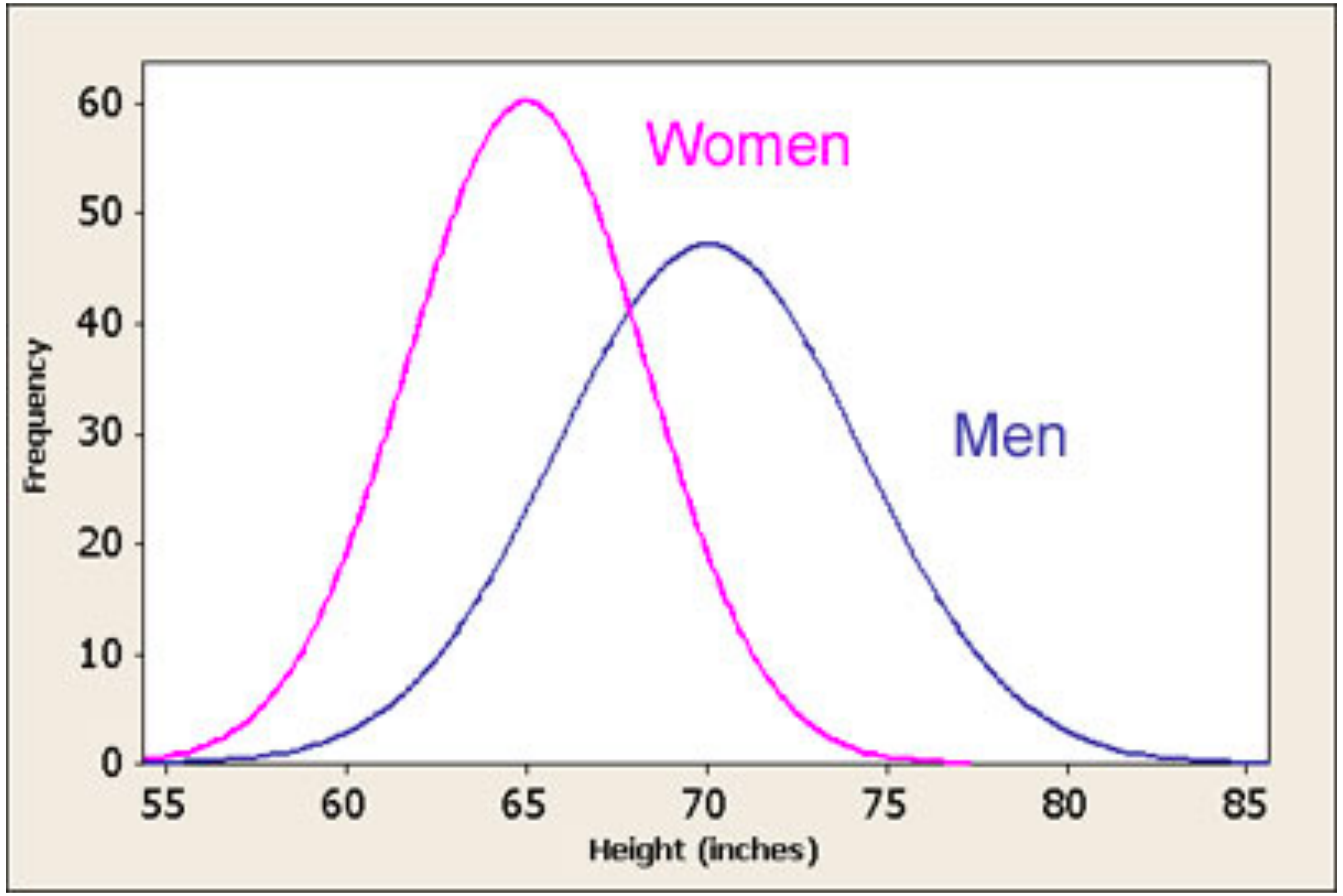
1. **Random Variables, Distributions**
2. Expectations and Standard Deviation
3. Law of Large Numbers
4. Randomized Assignment and Selection Bias
5. Assumptions of Causal Inference

Random Variables

- When studying experiments, we model outcomes as random.
- For example when flipping a coin, the outcome can be either heads or tails.
- Each flip of a coin is a random variable. It can take on two values but we don't know which will occur.
- A fair coin lands heads with probability 1/2 and lands tails with probability 1/2.

Random Variables (pt 2)

- Random variables take on many values and the probability of each value is determined by a function (called a probability density function).
- For example, a roll of the die takes each of the values {1, 2, 3, 4, 5, 6} with a value of 1/6.
- Some random variables, such as the height of a person (or the spending of a customer) are often modeled as continuous.
- In the case on the heights for men and women, this curve is the normal density function. Also known as the normal distribution.



Outline

1. Random Variables, Distributions
2. **Expectations and Standard Deviation**
3. Law of Large Numbers
4. Randomized Assignment and Selection Bias
5. Assumptions of Causal Inference

The Expectation of a Random Variable

- Suppose we flipped a fair coin and labeled heads as a 1 and tails as a 0.

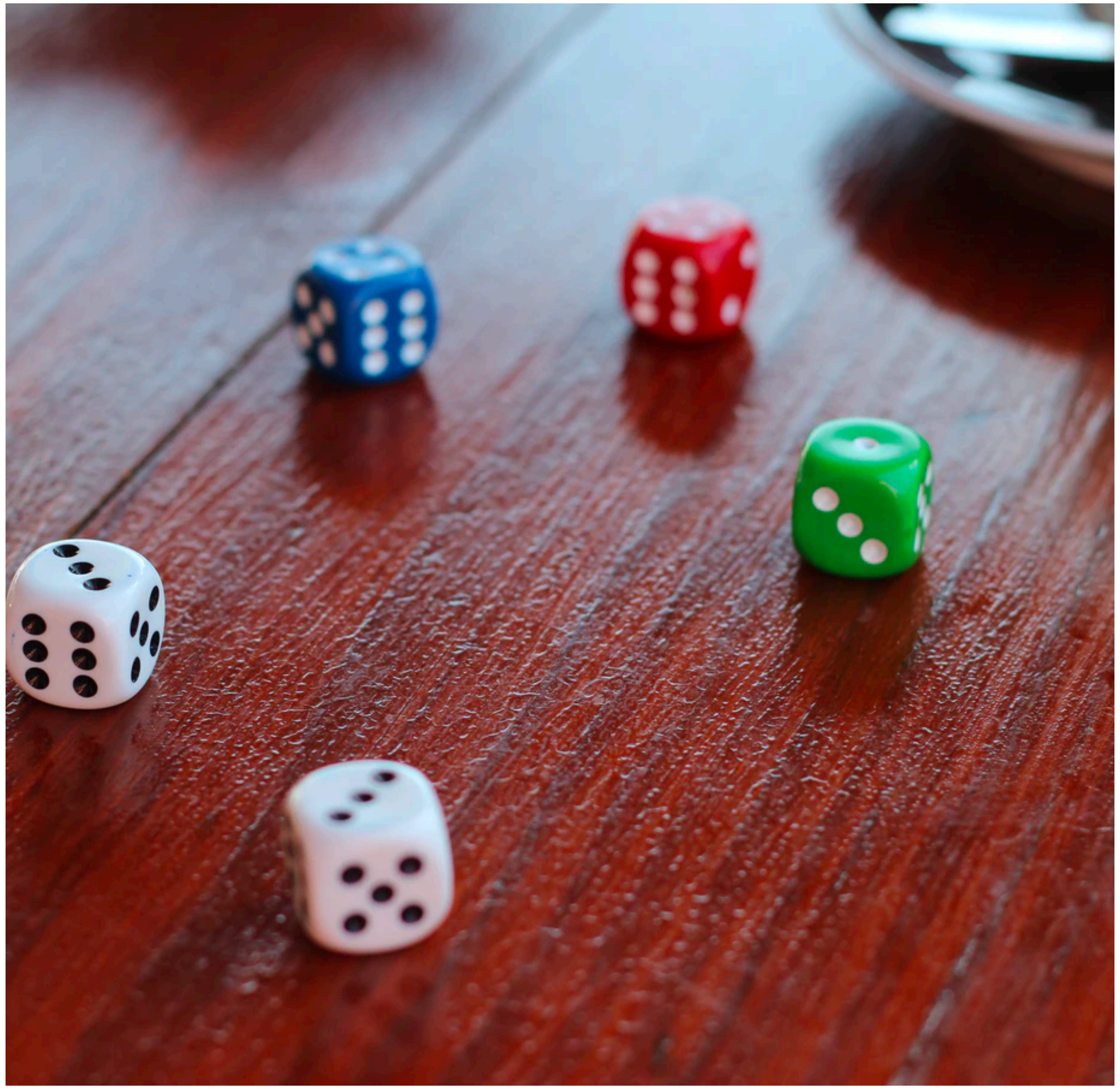
- The expected value would be equal to 1/2. $1 * 1/2 + 0 * 1/2 = 1/2$.

- The expectation of a random variable is the equal to each value that the random variable takes times the probability.

The Expectation of Dice Rolls

- Recall that a die takes on the values 1 through 6, and the probability of each value is 1/6.
- The expectation is:

$$
\begin{aligned}
&= 1 \cdot 1/6 + 2 \cdot 1/6 + \cdots + 6 \cdot 1/6 \\
&= 3.5
\end{aligned}
$$

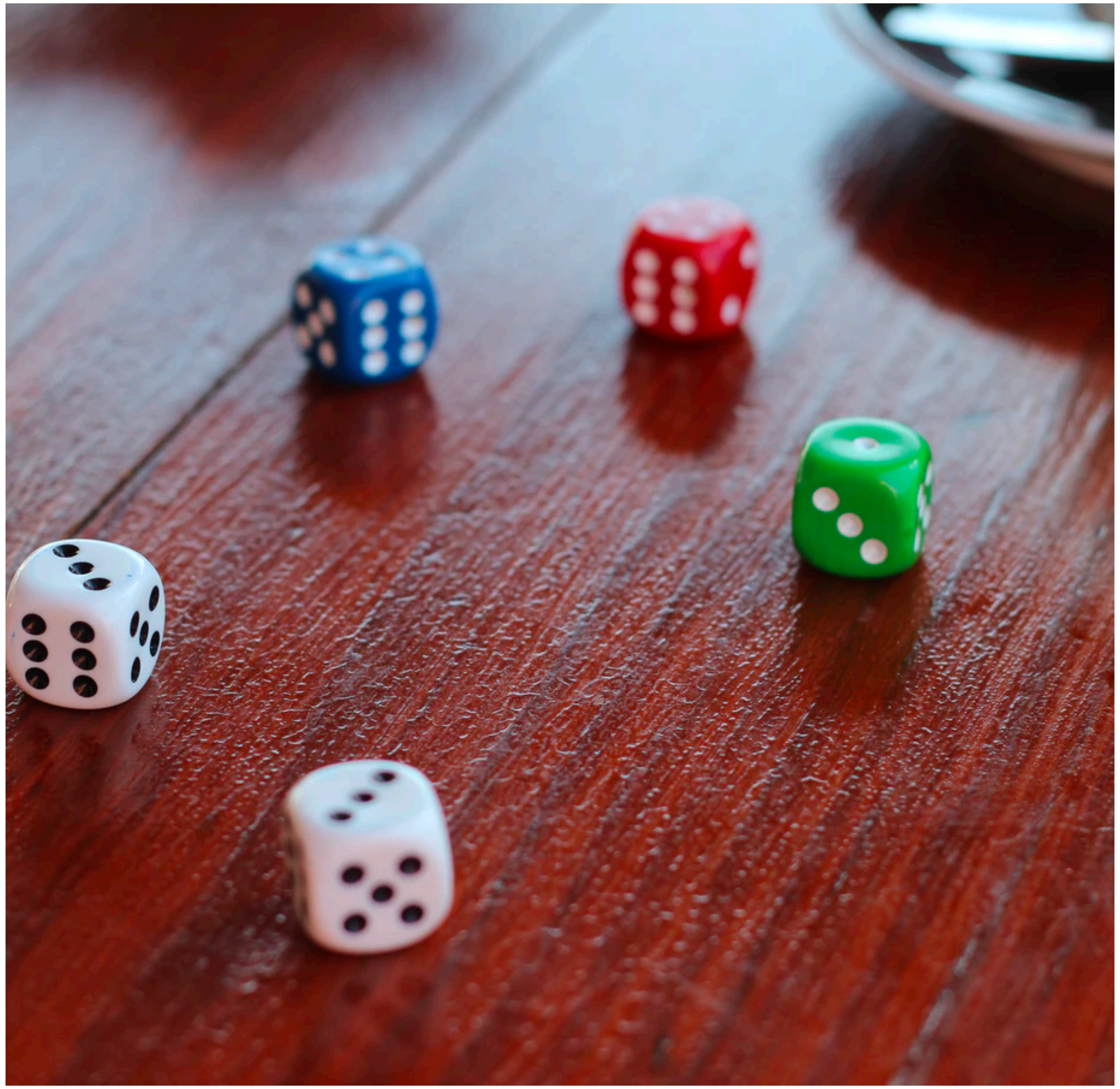


Notation

- Let X be a random variable and P(X = x) be the probability that X equals x.

- Then the expectation is denoted:

$$E[X] = \sum_{x} x \cdot P(X = x)$$



Properties of expectations

- The expectation of constant is the constant:

$$E[\alpha] = \alpha$$

- The expectation of random variable X and a constant:

$$E[\alpha + \beta X] = \alpha + \beta E[X]$$

- The expectation of the sum of two random variables.

$$E[X + Y] = E[X] + E[Y]$$

Our Example Causal Problem

- **Does attending class in-person (vs on Zoom) increase the probability of getting an A?**
- Let's say that the random variable is the individual treatment effect in our sample.
- -1/3 is the expectation of the treatment effect for the people in our experiment.

Person	In-person	Zoom	Treatment Effect
John	1	1	0
Mary	0	1	-1
Suraj	0	0	0
Katerina	1	1	0
Molly	0	1	-1
Leroy	0	0	0
Average	1/3	2/3	-1/3

Variance and standard deviations

- Variance is a measure of the spread of the distribution. The standard deviation is the square root of the variance.
- Intuitively, both measures how different the typical value is from its expectation.
- Formula:

$$\mu = \mathbb{E}[X]$$

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mu)^2\right]$$

Variance and standard deviations: Example

- Consider a deck of cards. Suppose that that the deck has cards 1, 2, 3, 4, 5, with equal probability (1/5).
- The expectation (average) of them is equal to 3.
- What about the variance?

$$2 = \frac{1}{5}(1-3)^2 + \frac{1}{5}(2-3)^2 + \frac{1}{5}(3-3)^2$$
$$+\frac{1}{5}(4-3)^2 + \frac{1}{5}(5-3)^2$$

Variance and standard deviations: Example

- Now let's pick a more extreme distribution. The deck has only 1s and 5s.

- The expectation is the same
  1/2 * 1 + 1/2 * 5 = 3.

- The variance is 4, which is bigger than 2 from the previous slide:

$$\frac{1}{2}(1-3)^2 + \frac{1}{2}(5-3)^2 = 4$$

- Intuitively, a deck with 1s and 5s is more spread out than one with 1, 2, 3, 4, 5 so it has a higher variance.

Outline

1. Random Variables, Distributions
2. Expectations and Standard Deviation
3. **Law of Large Numbers**
4. Randomized Assignment and Selection Bias
5. Assumptions of Causal Inference

The Law of Large Numbers

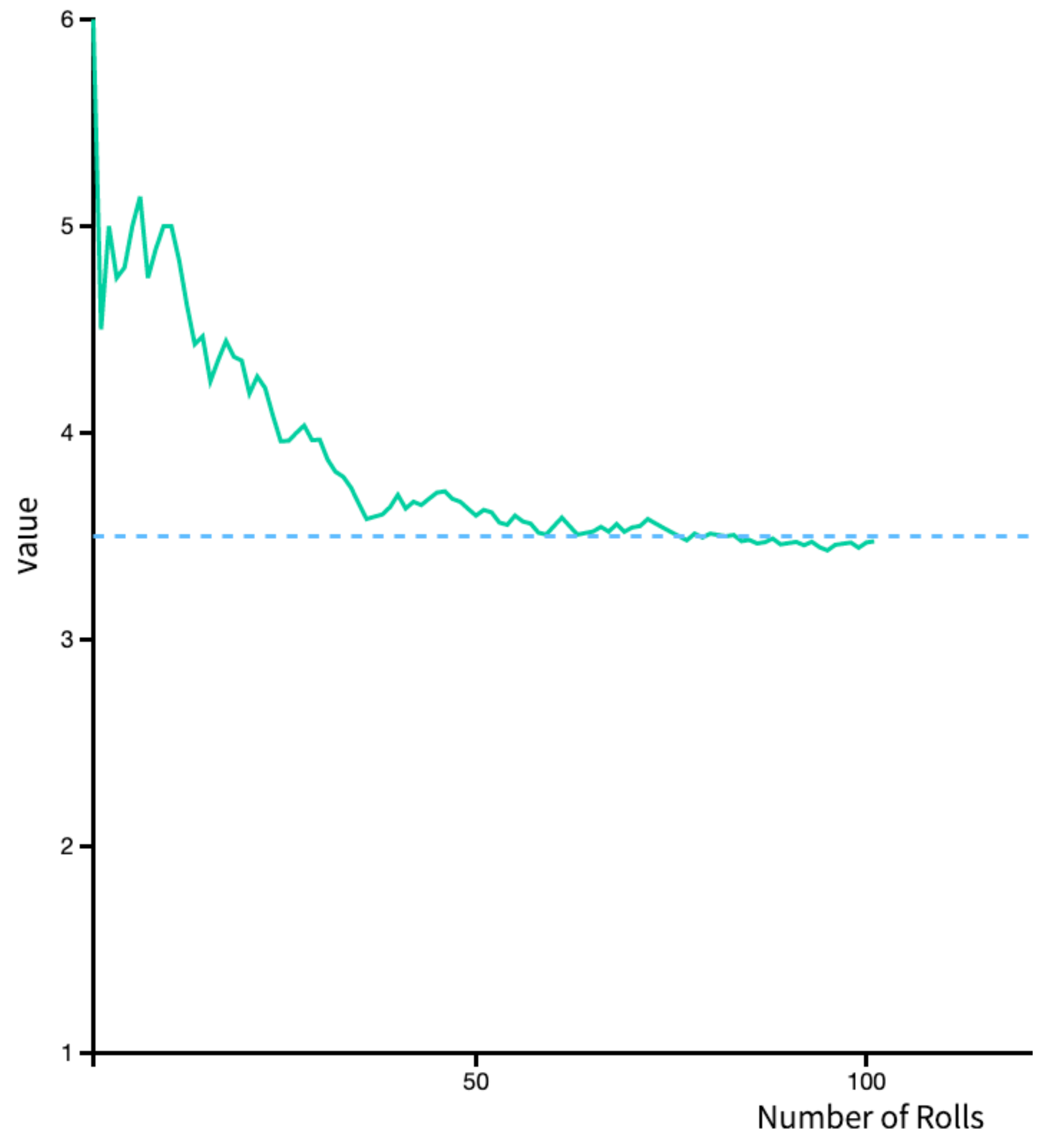
- Important concept, main intuition for why we want large samples.
- A sample is a set of draws of a random variable. For example, a roll of a die is a sample and 5 rolls of a die is a sample.
- The sample average, is simply the average of the sample. The sample average usually does not equal the expectation.
- For example, the expectation of a roll of the dice is 3.5. But if we rolled a dice one, it is impossible to get an average of 3.5 since the dice only takes the values 1, 2, 3, 4, 5, 6.
- **The law of large numbers: as the sample grows large, the sample average becomes close to the expectation.**

The Law of Large Numbers

- Imagine you roll a die and compute the sample average. Then roll it again and compute the sample average. And keep doing that for 100 times.

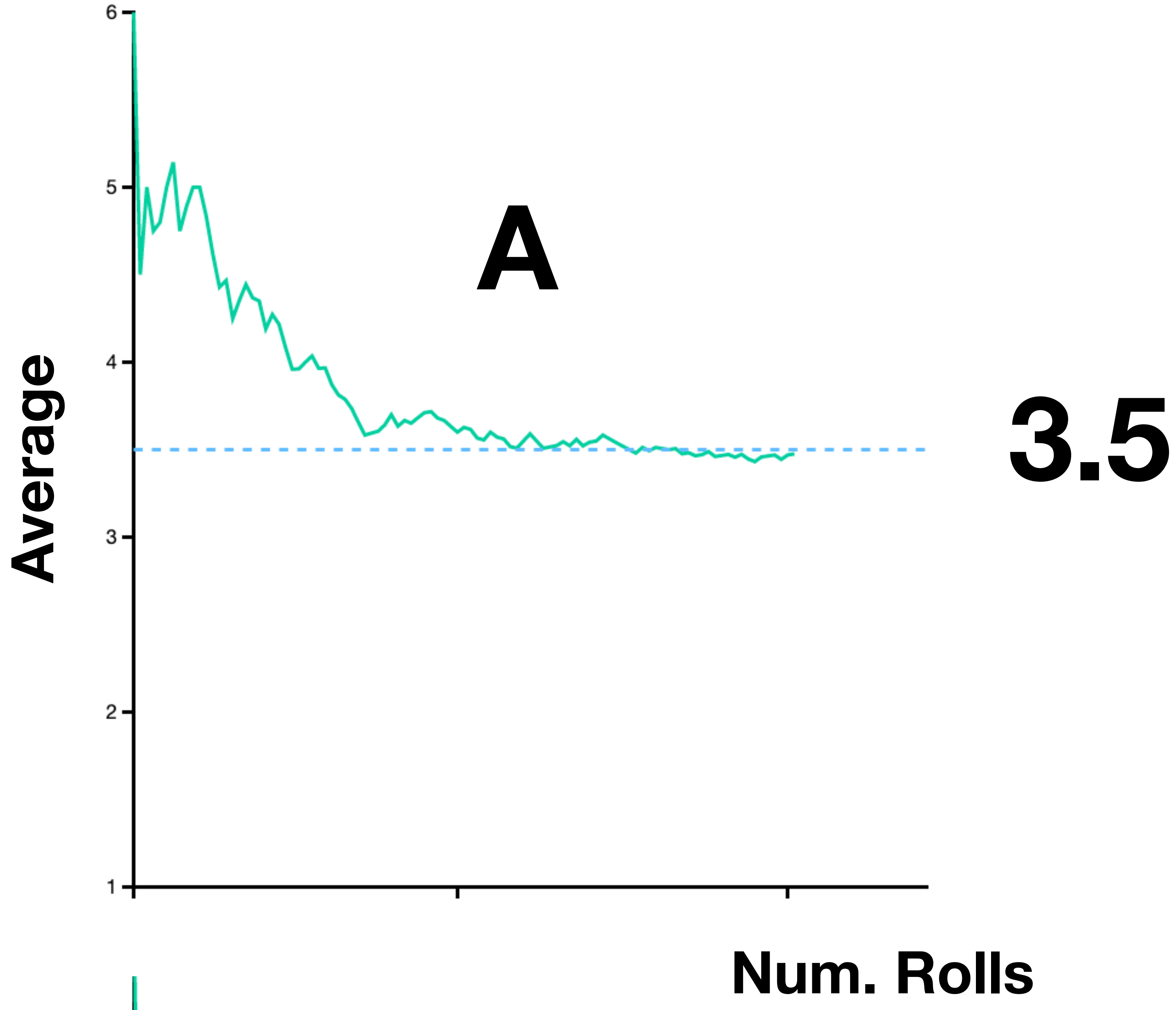
- The figure plots the sample average after each roll.

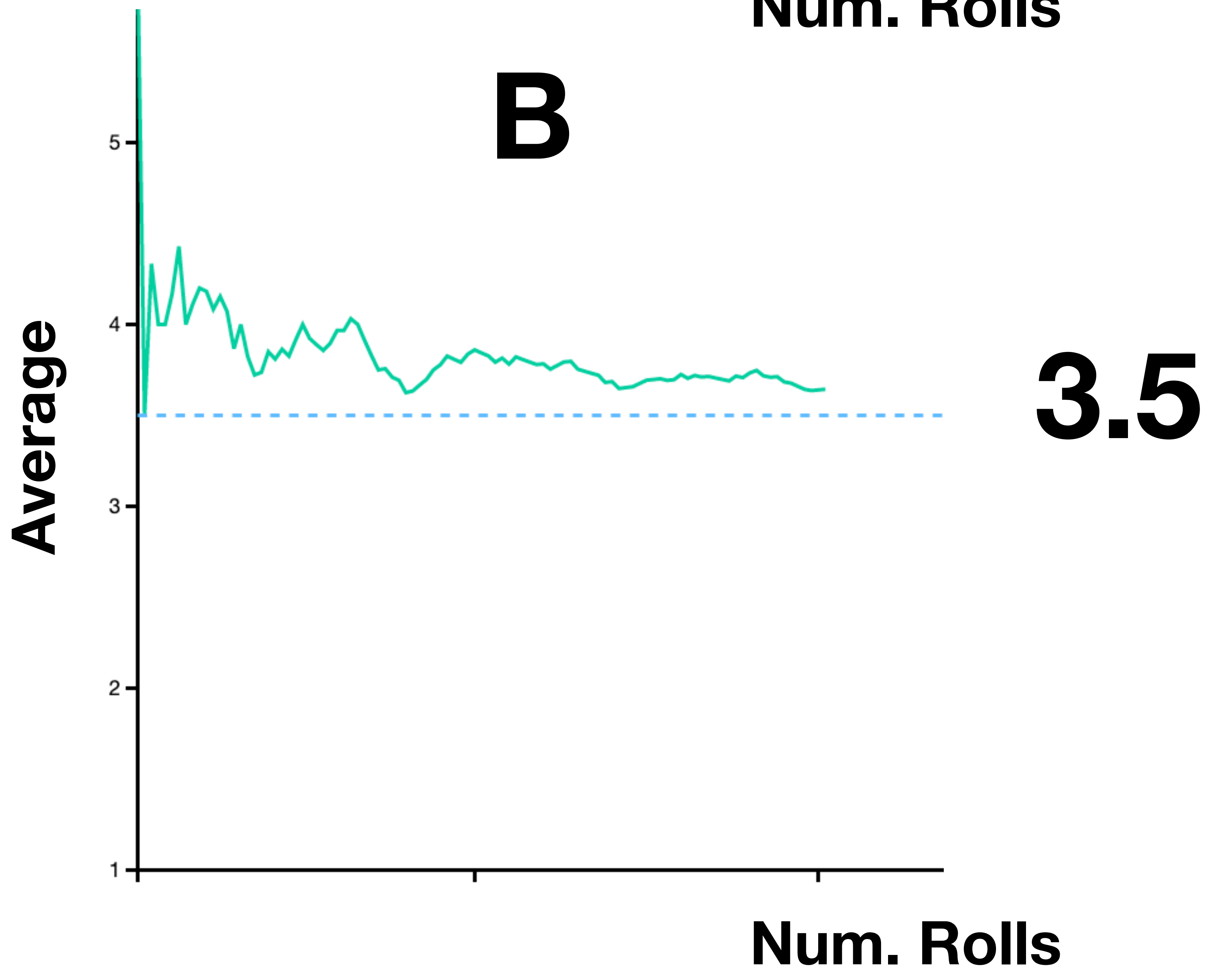
- You can see that the sample average approaches the expectation as the number of rolls increases.



The Law of Large Numbers

- Imagine you roll a die and compute the sample average. Then roll it again and compute the sample average. And keep doing that for 100 times.
- Figure A plots the sample average after each roll.
- You can see that the sample average approaches the expectation as the number of rolls increases.
- We can do this procedure again, figure B, and we will get a different path. But it still approaches the expectation.





Interlude

https://seeing-theory.brown.edu/

Outline

1. Random Variables, Distributions
2. Expectations and Standard Deviation
3. Law of Large Numbers
4. **Randomized Assignment and Selection Bias**
5. Assumptions of Causal Inference

Let’s go back to our example: In-person vs Zoom

Person	In-person	Zoom	Treatment Effect
John	1	1	0
Mary	0	1	-1
Suraj	0	0	0
Katerina	1	1	0
Molly	0	1	-1
Leroy	0	0	0
Average	1/3	2/3	-1/3

Let's randomly assign people to 'In-person'.

- **Simple** (e.g. toss coin for every subject)
  - Can create wrong proportions when sample is small. By chance, 1 person could be in the treatment and 5 in the control.
- **Complete** randomization. Suppose we want 3 treated and 3 control individuals.
  - **Randomly** order all participants and let the first 3 be treated.

Person	In-person	Zoom	Treatment Effect
John	1	1	0
Mary	0	1	-1
Suraj	0	0	0
Katerina	1	1	0
Molly	0	1	-1
Leroy	0	0	0
Average	1/3	2/3	-1/3

These are not randomized!!!

Every other day

Day of week

Alternation

Using modulus arithmetic

Notation:

$Y_i(1)$ ← Potential outcome for person 'i' if that person is treated.

$Y_i(0)$ ← Potential outcome for person 'i' if that person is NOT treated.

	Outcome In-person	**Outcome Zoom**	
Person (i)	$Y_i(1)$	$Y_i(0)$	Effect
John	1	1	0
Mary	0	1	-1
Suraj	0	0	0

Let’s randomly assign people to ‘In-person’.

- Bold, colored numbers are the one we would observe in an experiment.
- Let’s call the average difference between treated (purple) and control (orange) individuals the estimate of the average treatment effect (or $\widehat{ATE}$).
- We put the ‘hat’ on quantities that estimated
  - This equals .5 - .5 = 0.
- Note, $\widehat{ATE}$ does not equal the true average treatment effect (ATE).

Treated?	Person	Y in Data	$Y_i(1)$	$Y_i(0)$	Effect
1	John	**1**	**1**	1	0
0	Mary	**1**	0	**1**	-1
0	Suraj	**0**	0	**0**	0
0	Katerina	**1**	1	**1**	0
1	Molly	**0**	**0**	1	-1
0	Leroy	**0**	0	**0**	0
	Average		1/3	2/3	-1/3

Let's randomly assign people to 'In-person'.

- Bold, colored numbers are the one we would observe in an experiment.
- Let's call the average difference between treated (purple) and control (orange) individuals the estimate of the average treatment effect (or $\widehat{ATE}$).
- We put the 'hat' on quantities that estimated
  - This equals .5 - .5 = 0.
- Note, $\widehat{ATE}$ does not equal the true average treatment effect (ATE).

Treated?	Person	Y in Data	$Y_i(1)$	$Y_i(0)$	Effect
1	John	**1**	**1**	1	0
0	Mary	**1**	0	**1**	-1
0	Suraj	**0**	0	**0**	0
0	Katerina	**1**	1	**1**	0
1	Molly	**0**	**0**	1	-1
0	Leroy	**0**	0	**0**	0
	Average		1/3	2/3	-1/3

Why randomize treatment?

- Randomization ensures that the process by which treatment is assigned is not related to anything else about the person.
    - For example, the chance that a motivated student is treated is equal to the chance that a non-motivated student is treated.
- When the number of participants is small, it could be that by chance, many more motivated students are treated.
- Because of the law of large numbers, as the number of participants increases, the estimated average treatment effect gets close to the true average treatment effect.

Estimators vs Estimates

- Estimators are procedures for taking our data and producing a number. That number is called an estimate.
  - The procedure of taking the average difference between treatment outcomes and control group is an estimator. We will discuss others in class.

Estimators vs Estimates

- What do we want from the estimator?
  - We want estimators to produce estimates that are close to the true parameter we're interested in (such as the ATE).
  - An estimator is called 'unbiased' for a parameter if its expectation is equal to the true value of the parameter.
  - The difference in means in an experiment is an **unbiased estimator** of the ATE.

External Validity

Selection Bias vs Non-representative sample.

- Let’s say 1000 people volunteered to participate in a vaccine trial. We randomly assign 500 to the vaccine and the others to a placebo (we will discuss this later). We do not have to worry about selection bias.
- There may be, however, be a non-representative sample.

Selection Bias vs Non-representative sample.

- If only young people volunteer in the trial. We do not learn about the treatment effects for old people.
- This problem is often called these names:
  - Non-representative sample
  - External validity
  - Generalizability
- **You need to understand the difference between non-representative samples and selection bias! (We will return to this many times in the course)**

(Optional) Proof of why randomization eliminates selection bias as the sample grows large.

$$E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0] = E[Y_i(1)] - E[Y_i(0)] = ATE$$

- The law of large numbers ensures that the sample average outcome of the treated will be close to the expectation:

$$E[Y_i(1) \mid D_i = 1]$$

- Similarly, the law of large numbers ensures that the sample average outcome of the control will be close to the expectation:

$$E[Y_i(0) \mid D_i = 0]$$

- Randomization allows us to get rid of the conditional.
- The above equation shows that the difference in these is the ATE, which is what we'd like to measure.

Outline

1. Random Variables, Distributions
2. Expectations and Standard Deviation
3. Law of Large Numbers
4. Randomized Assignment and Selection Bias
5. **Assumptions of Causal Inference**

Assumption 1: “Excludability”

- The treatment is the only thing that changes between treatment and control.
- Example where this fails: a drug study with no placebo drug.
- Person in the treatment group takes a drug, and thinks he is taking a drug.
- ‘Placebo effect’ -> even taking a ‘fake’ drug like a sugar pill improves outcomes.
- Therefore, we want the person in the control group to take a ‘placebo’.

Assumption 2: “Non-interference”

- The treatment of one individual in the experiment does not affect the outcome of other individuals in the experiment.
- This can fail when individuals interact. If we randomize students in a class to in-person vs zoom we may have a violation.
- In-person students interact with zoom students and vice versa. Their learning can affect the control group’s learning.

Assumption 2: “Non-interference”

Also called:
“no spillovers”
“stable unit treatment value assumption” (SUTVA)

Recap

1. Random Variables, Distributions
2. Expectations and Standard Deviation
3. Law of Large Numbers
4. Randomized Assignment and Selection Bias
5. Important Assumptions

Why actually happened?

IZA DP No. 14356

Zooming to Class?: Experimental Evidence on College Students' Online Learning during COVID-19

Michael S. Kofoed
Lucas Gebhart
Dallas Gilmore
Ryan Moschitto

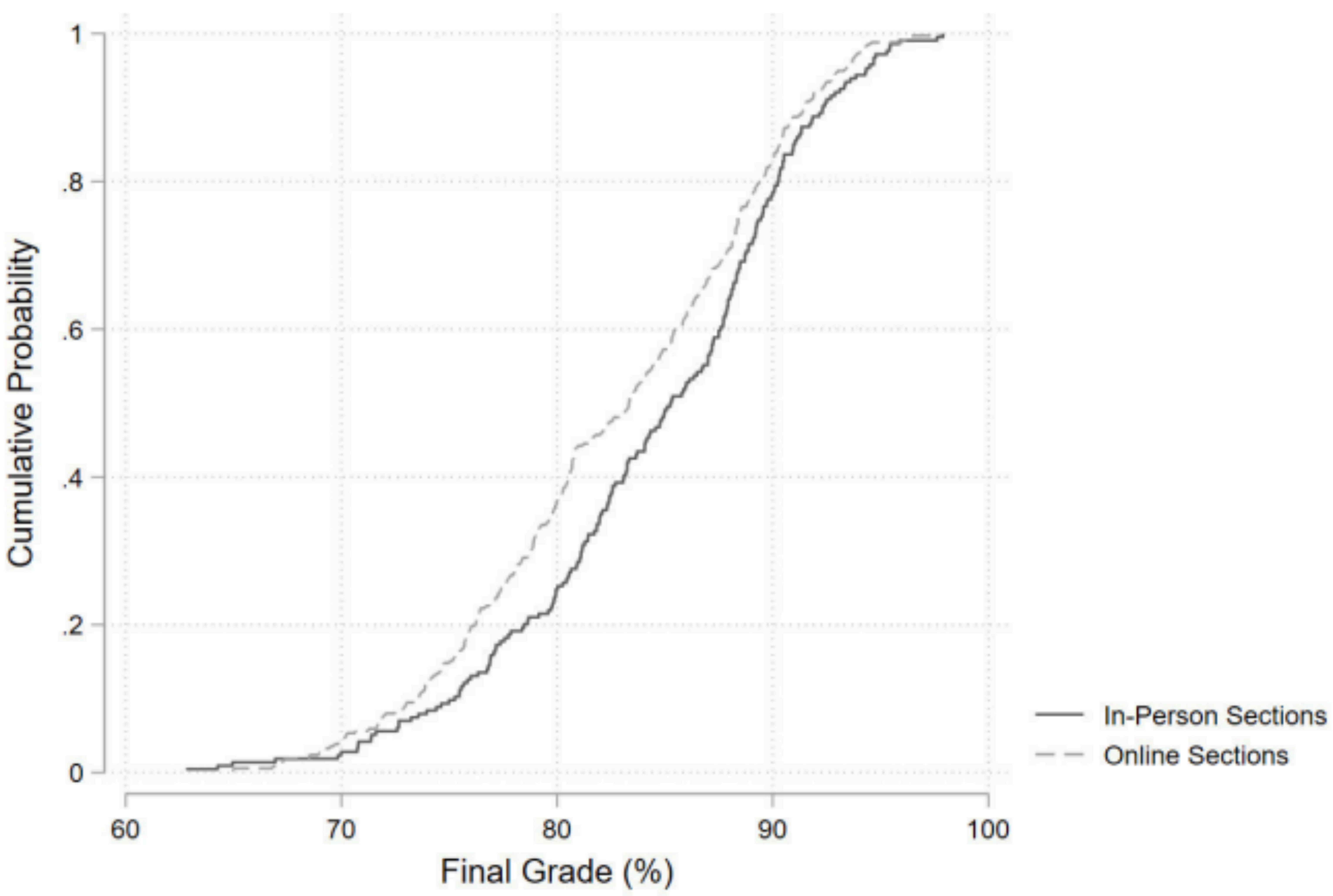
Setting

- West Point Military Academy
- Everyone takes the same classes, so there are many sections.
- Due to social distancing requirements during COVID, in-person classrooms couldn’t fit everyone.
- Each in-person and online class allowed for 12 and 18 students respectively.
- Students randomized into section (337 in online and 214 in person).

Result

- Students online earned .234 standard deviations lower grade, around half a +/- grade.

Figure 2: Cumulative Distributive Function for Final Course Grade by Teaching Modality



Next Time: Measuring Uncertainty